The two main jobs of a computer are I/O and computing. In many cases, the main job is I/O, and the computing or processing is merely incidental. For instance, when we browse a web page or edit a file, our immediate interest is to read or enter some information, not to compute an answer.

The role of the operating system in computer I/O is to manage and control I/O operations and I/O devices. Although related topics appear in other chapters, here we bring together the pieces to paint a complete picture of I/O. First, we describe the basics of I/O hardware, because the nature of the hardware interface places constraints on the internal facilities of the operating system. Next, we discuss the I/O services provided by the operating system and the embodiment of these services in the application I/O interface. Then, we explain how the operating system bridges the gap between the hardware interface and the application interface. We also discuss the UNIX System V STREAMS mechanism, which enables an application to assemble pipelines of driver code dynamically. Finally, we discuss the performance aspects of I/O and the principles of operating-system design that improve I/O performance.

Bibliographical Notes

[Vahalia (1996)] provides an overview of I/O and networking in UNIX. [McKusick and Neville-Neil (2005)] detail the I/O structures and methods employed in FreeBSD. The use and programming of the various interprocess-communication and network protocols in UNIX are explored in [Stevens (1992)]. [Hart (2005)] covers Windows programming.

[Intel (2011)] is a good source of information for Intel processors. For a discussion of STREAMS, see [Rago (1993)]. [Hennessy and Patterson (2012)] describe multiprocessor systems and cache-consistency issues.

Bibliography

Chapter 12  I/O Systems


